

Paranoid Secondary: Waterfilling in a Cognitive Interference Channel with Partial Knowledge

Debashis Dash and Ashutosh Sabharwal

Abstract

We study a two-user cognitive channel, where the primary flow is sporadic, cannot be re-designed and operating below its link capacity. To study the impact of primary traffic uncertainty, we propose a *block activity model* that captures the random on-off periods of primary's transmissions. Each block in the model can be split into parallel Gaussian-mixture channels, such that each channel resembles a multiple user channel (MAC) from the point of view of the secondary user. The secondary senses the current state of the primary at the start of each block. We show that the optimal power transmitted depends on the sensed state and the optimal power profile is *paranoid*, i.e. either growing or decaying in power as a function of time. We show that such a scheme achieves capacity when there is no noise in the sensing. The optimal transmission for the secondary performs rate splitting and follows a *layered water-filling* power allocation for each parallel channel to achieve capacity. The secondary rate approaches a genie-aided scheme for large block-lengths. Additionally, if the fraction of time primary uses the channel tends to one, the paranoid scheme and the genie-aided upper bound get arbitrarily close to a no-sensing scheme.

Index Terms

Cognitive radio, spectrum sensing, interference channel, Gaussian mixture channel, capacity, side information, rate splitting, water-filling.

I. INTRODUCTION

Cognitive wireless is a novel approach to deploy new wireless services in the presence of legacy devices with priority access to the channel. The aim of a cognitive framework is to

The authors are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA. (email: {ddash, ashu}@rice.edu). This work was partially supported by NSF Awards CCF-0635331 and CNS-0325971.

communicate as an underlay to an underutilized primary network without degrading the primary communication beyond a predetermined threshold. The key constraint for our formulation is that the primary transceiver is legacy but fixed i.e. it was not designed keeping a cognitive secondary in mind. This constraint makes the classical interference channel [1] one sided, i.e, Z-interference channel with an additional power constraint. In this paper, we analyze how temporal opportunities due to sporadic primary traffic can be exploited, even if the secondary transmitter has incomplete information about the current state of primary traffic.

Our contributions are three-fold. First, we approximate the uncertainty in primary activity (interference channel with a sporadic primary) by a simple block activity model where the primary changes its state at most once during a block of fixed duration $T + 1$. In this model, the two sources of uncertainty in the primary traffic are captured by the initial state s_0 and the time of state change τ . Their actual values are unknown to the secondary, but their distribution $\pi(s_0)$ and $f_T(\tau)$ are known. We first show that the reliability constraint at the primary receiver, which requires that primary transmission should not be harmed, places an additional power constraint on the secondary transmitter (see for example, [25]) and the interference caused by the sporadic primary at the secondary receiver converts the AWGN channel to a Gaussian mixture channel.

Second, we present two sense-and-send schemes where the secondary senses the channel at the start of each block to look for temporal opportunities. The fixed primary design converts the effective channel into a MAC. The secondary splits its rate into two layers [20], [19]. We show that it is optimal to treat the primary message either as all public information or as all private information but not both because the fixed primary's message cannot be split into private and public parts. We also present a no-sensing scheme and a genie-aided upper bound. We prove that for the sense-and-send schemes the secondary power profile is paranoid, i.e. the power monotonically increases or decreases during a block depending on whether the primary was using the channel or not during the start of the block. The paranoid profile arises due to the effective noise of the mixture channels which decays or grows, depending on the starting state of the channel in a block. Additionally, we prove that when the sensing is perfect, the paranoid scheme is optimal. The proof is along the lines of a channel with delayed state information [22].

Third, we show that if there is no information about the primary traffic the rate splitting scheme achieves capacity. The effective channel is converted to a MAC with a fixed transmitter from the secondary's point of view and hence rate splitting at the secondary transmitter and

sequential decoding at the secondary receiver is optimal. Finally, we show that the paranoid scheme approaches a genie-aided upper bound for large block lengths and when the primary traffic becomes more persistent, both these schemes get close to the no-information scheme.

Much like any active area, many variations have been proposed and studied in the literature [12, and references therein]. Some of the earlier work [16, 18] in opportunistic spectrum allocation for cognitive flows was motivated by studies done by FCC [7] showing significant spectral underutilization. Many aspects of cognitive radio have been studied including coding with degraded messages [5, 23], capacity of the secondary flow with causal or non-causal information [15], stability of the queues at both the flows for maximal secondary rate with a guaranteed primary throughput [21], spectral shaping [25] etc. The case of an adaptive primary which acts as a feedback to the cognitive radio is considered in [6]. When interference is considered in terms of SINR constraints (see for example, [4, 9–11]), the effect of the primary communication at the secondary is not considered. Power control schemes have been derived for different power constraints [24, 25] when the primary is persistent but operating below capacity. The idea of opportunistic interference cancellation was introduced in [19]. However all the above results do not address the issue of the uncertainty in the channel when the primary traffic is fixed but sporadic. We extend the idea of opportunistic interference cancellation to derive sense and send schemes for a fixed and sporadic primary operating below capacity.

The rest of the paper is organized as follows. In Section II, we present the system model and introduce the a block activity model for a sporadic primary. In Section III, we first present a genie-aided bound. Then we derive the paranoid secondary profile for a sense and send protocol when the sensing is noisy. Finally, we give two special cases when the sensing is noiseless and when no sensing is performed. We conclude with some numerical examples.

II. SYSTEM MODEL

A. Signal Model

We consider an interference channel (Figure 1), where the channel inputs and outputs are related as, $Y_1 = h_{11}X'_1 + h_{21}X'_2 + Z_1$ and $Y_2 = h_{12}X'_1 + h_{22}X'_2 + Z_2$, where X'_1, X'_2 are channel inputs, Y_1, Y_2 are the channel outputs and Z_1, Z_2 are independent zero mean, unit variance Gaussian noise. Each transmitter is individually average power constrained, such that $E[X'^2_1] \leq P_1$ and $E[X'^2_2] \leq P_2$. We define the following for convenient interpretation: $\text{SNR}_1 = |h_{11}|^2 P_1$, $\text{SNR}_2 =$

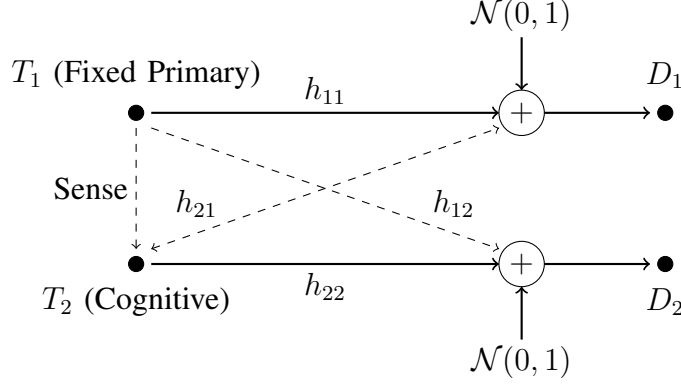


Fig. 1. A two-flow cognitive interference network.

$|h_{22}|^2 E[X_2'^2]$, $\text{INR}_1 = |h_{21}|^2 E[X_2'^2]$, $\text{INR}_2 = |h_{12}|^2 P_1$. We have fixed the primary power to P_1 (primary uses its full power budget) in our definition of SNR_1 and INR_2 , leaving the secondary power as an optimization variable in our definition of SNR_2 and INR_1 . Finally, we use the standard form [2] of interference channel by assuming $|h_{11}| = |h_{22}| = 1$. Thus the channel input-output are related as $Y_1 = \sqrt{\text{SNR}_1}X_1 + \sqrt{\text{INR}_1}X_2 + Z_1$ and $Y_2 = \sqrt{\text{INR}_2}X_1 + \sqrt{\text{SNR}_2}X_2 + Z_2$, where the variance of X_1 and X_2 is bounded above by one.

B. Block Activity Traffic Model for the Primary Flow

The primary is assumed to have the following two properties: its channel is *underutilized* i.e. it sends below the capacity of its own channel when it transmits, and its traffic is *sporadic* i.e. it does not occupy the channel continuously. We model the channel underutilization by assuming that the primary employs Gaussian random codes with a code rate of $\mathcal{C}\left(\frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}}\right)$, where $\text{INR}_{\text{gap}} \geq 0$ and $\mathcal{C}(x) = \frac{1}{2} \log_2(1+x)$. The sporadic traffic is modeled by assuming that the primary transmits for β fraction of time, which results in an average primary rate of $R_1 = \beta \mathcal{C}\left(\frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}}\right)$. Thus a secondary user can use the same channel as long as it ensures that the primary INR does not go above INR_{gap} . We refer to this as the *INR constraint*. We capture the two uncertainties from the secondary's point of view, the start time of the primary transmission bursts and the duration of the bursts, using the following block activity model. The primary transmissions are assumed to occur in blocks of $T+1$ time-slots, where T is a finite constant. All transmissions by the primary and the secondary are considered to be slot synchronous. The

primary user activity is labeled as the state of the primary channel and is denoted by s_t , where $0 \leq t \leq T$ is the time-slot index in a block. The state process is independent of the secondary channel's input and output. The primary channel is either in the busy state, $s_t = 1$, or in the idle state, $s_t = 0$ during the time-slot t . To keep the analysis tractable, we assume that the primary user changes its state only once at time-slot τ in a block (see Figure 2). Thus, the tuple (s_0, τ) captures the two uncertainties related to the sporadic transmissions.

The starting state of the primary user for each block is assumed to be independently drawn from $\pi(s_0) = [\pi(0), \pi(1)]$. Conditioned on the starting state, the switching time of the primary τ has a probability mass function, $f_T(\tau), 1 \leq \tau \leq T + 1$ and the corresponding distribution function $F_T(\tau) = \sum_{i=1}^{\tau} f_T(i)$. If $\tau = T + 1$, then the primary user does not change its state during the block. The secondary transmitter is assumed to know $\text{INR}_{\text{gap}}, \pi(s_0)$ and $f_T(t)$. The secondary receiver is assumed to have perfect knowledge to do coherent decoding.

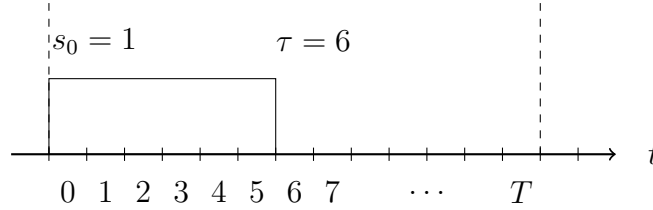


Fig. 2. Simplified block activity model for the primary packets with only one allowed switching of state. This model is characterized by the starting state s_0 , the state-switching time τ and the blocklength $T + 1$.

Since the primary uses Gaussian codes, from the point of view of the secondary the channel always behaves like an AWGN channel. The effective noise as seen by the secondary receiver has the mixture distribution for $1 \leq t \leq T$,

$$Z^{(t)} = \begin{cases} \mathcal{N}(0, 1) & \text{if } s_t = 0 \text{ (probability } = \bar{\beta}_s(t)) \\ \mathcal{N}(0, 1 + \text{INR}_2) & \text{if } s_t = 1 \text{ (probability } = \beta_s(t)) \end{cases}, \quad (1)$$

where, $\bar{\beta}_s(t) = 1 - \beta_s(t)$. The term $\beta_s(t)$ is the probability of the state $s_t = 1$ conditioned on the starting state s , and is used to derive the power profile in Section III. Note that the β is the actual primary channel usage fraction and $\beta_s(t)$ is what is seen by the secondary. A persistent primary can be represented by putting $\beta = 1$. If $\text{INR}_{\text{gap}} = 0$, the system is fully loaded and no secondary is allowed on the same channel. The interesting case is when $\text{INR}_{\text{gap}} > 0$ and $\beta < 1$.

C. Interleaved Block Code for the Secondary Flow

We now define the set of secondary codes over B consecutive blocks, when the secondary has an estimate of the starting state \hat{s}_0 for each block. Bold face letters represent the vector corresponding to the variable for all blocks, e.g. $\mathbf{s}_0 = (s_0^{(1)}, s_0^{(2)} \dots, s_0^{(B)})$ represents the starting state of all B blocks. Since the t^{th} time-slot of each block has identically distributed noise as given by Equation (1), the secondary designs T different codebooks each matched to a given time slot across B blocks. For example, during b^{th} block the secondary transmits $(x_1^{(b)}, x_2^{(b)}, \dots, x_T^{(b)})$ where $x_j^{(i)}$ denotes the i^{th} component of a codeword from the j^{th} codebook. Similarly the t^{th} codeword is given by $\mathbf{x}_t = (x_t^{(1)}, x_t^{(2)} \dots, x_t^{(B)})$ as shown in Figure 3.

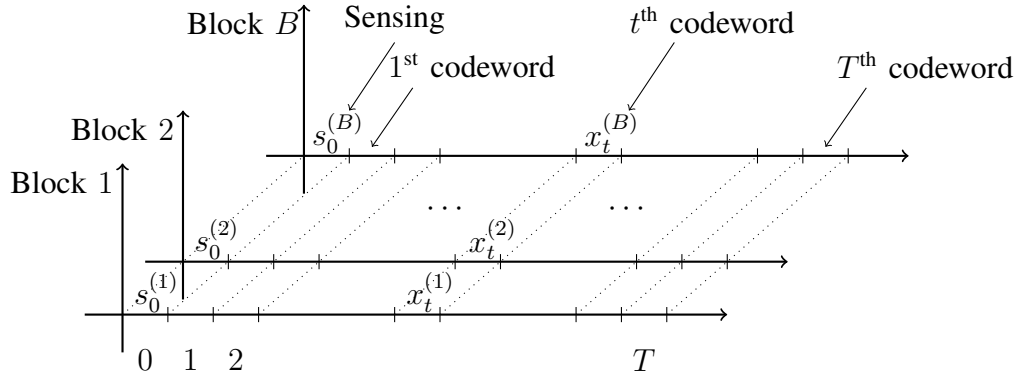


Fig. 3. Codewords are multiplexed across B blocks, i.e. each time slot uses a different codeword which spans across B blocks. The figure shows the sensing operation at the beginning of each block and the t^{th} codeword, $(x_t^{(1)}, \dots, x_t^{(B)})$. Each of the T codewords are picked from the corresponding T codebooks. The block size T is a positive constant, however, the code length B tends to infinity.

A cognitive interleaved block code of length $N = BT$ consists of T separate $(B, 2^{B(R-\epsilon)}, \epsilon)$ -codes, corresponding to each time slot of a block. Each of these codebooks further consists of two component codebooks corresponding to the sensed starting state of the block \hat{s}_0 which may or may not match the actual starting state s_0 . The codebooks are defined by a set of $2T$ encoding functions that map the set of equiprobable messages $\{1, \dots, 2^{BR}\}$ to channel inputs. The transmitter encoding function is defined as $f_{t\hat{s}_0} : \mathcal{W} \times \mathcal{S} \rightarrow \mathcal{X}$ for $t = 1, \dots, T$ and $\hat{s}_0 = 1, 2$, i.e. $\mathbf{x}_{t\hat{s}_0} = f_{t\hat{s}_0}(w, \hat{s}_0)$. The decoding function $g : \mathcal{Y}^B \times \mathcal{S}^B \rightarrow \mathcal{W}$ maps the received vector to the message set, leading to probability of error $P_e = \frac{1}{2^{BR}} \sum_{i=1}^{2^{BR}} \sum_{\mathbf{s}_0} p(\mathbf{s}_0) \sum_{\mathbf{y}: g(\mathbf{y}, \mathbf{s}_0) \neq i} P_e(\mathbf{y} | \mathbf{f}(i), \mathbf{s}_0)$. The rate R is ϵ -achievable if there exists a cognitive block code for sufficiently large $N = BT$ and

$P_e \leq \epsilon$. A rate R is said to be achievable if there exists an ϵ -achievable code for every $\epsilon > 0$ and the capacity is defined to be the supremum of all the achievable rates R .

Finally consider the Han-Kobayashi scheme for interference channels where each user's data is split into private and public parts. The public information is decodable by both decoders and the private information is decodable by the intended decoder. A special case of this structure is usable in our model. Firstly, since the primary encoder is fixed, its data can be considered either public (if the secondary can decode it, i.e. the cross channel can support the rate) or private, but it cannot be split into both. Secondly, since the primary decoder is fixed, it considers all the secondary information as noise, i.e. private. Since the primary data is fixed, the secondary can do rate splitting at the encoder and sequential decoding at the decoder without losing optimality. We will call the first layer the single user codeword with power ρ_{sNt} since it is decoded using a single user decoder (treating everything else as *Noise*), and the second layer the multiuser codeword with power ρ_{sSt} since it is decoded using a multiuser receiver (with *Successive* interference cancellation). The subscript s, \hat{s} will be replaced by its value (0 or 1) depending on the context.

III. MAIN RESULTS

In the block activity model, the two sources of uncertainty in the primary traffic are captured by the initial state s_0 and the time of state change τ . Their actual values are unknown to the secondary, but their distribution $f_T(\tau)$ and $\pi(s_0)$ are assumed to be known. Our aim is to understand how lack of knowledge of these two parameters impacts the secondary rate. We first derive the case where a Genie provides the information about (s_0, τ) which serves as an upper bound and a design motivation for the general case of Section III-B, where the secondary has an estimate \hat{s}_0 of the initial state s_0 and has no knowledge of τ . We then consider two important special cases of the general scheme. First, we consider the case of perfect sensing, $\hat{s}_0 = s_0$, and derive stronger results about the optimal secondary transmission design. Second, we consider the case where the secondary has no information about any of the unknown parameters, which is equivalent to \hat{s}_0 not providing any useful information about s_0 .

A. Genie-aided Case: Secondary has perfect estimate of (s_0, τ)

In this section, we assume Genie-aided knowledge of (s_0, τ) at the secondary. As described in Section II-B, the primary flow has two states, on and off. Hence the secondary transmitter uses

two interleaved codes matched to the two channel states. Such a solution is called water-filling or water-pouring [8], as the variance of the two codebooks is inversely proportional to the noise variance of the corresponding channel state. We use the following result from [17] to show that Gaussian codebooks are optimal for the secondary.

Lemma 1 (Optimality of Gaussian codewords, see [17, Theorem 1]): For a single-user scalar additive noise channel with nearest neighbor decoding irrespective of the noise distribution, the average probability of error over the ensemble of Gaussian codebooks of power P , approaches zero as the blocklength n tends to infinity for code rates below $\log(1 + P/N)$ (and approaches one for rates above $\log(1 + P/N)$).

Since the primary uses a Gaussian codebook, the equivalent noise at the secondary is also Gaussian. Further since the decoder of the primary is fixed, any distribution of noise at its receiver does not change the primary capacity as long as the variance (interference plus noise) is below INR_{gap} . Hence Gaussian codebooks are optimal for the secondary. The secondary uses a Gaussian codebook with power ρ_0 when the primary is off, and a superposition Gaussian codebook with power (ρ_{1N}, ρ_{1S}) . Define $\text{INR}_C = \frac{\text{INR}_{\text{gap}}}{|h_{21}|^2}$ and $\text{SIC}_C = |h_{12}|^2(\text{INR}_{\text{gap}} + 1) - 1$ to characterize the power allocation for the Gaussian codebook that achieves capacity.

Theorem 1 (Genie-aided Upper Bound): When $\text{SIC}_C \geq \min(\text{SNR}_2, \text{INR}_C)$, then the optimal power allocation $(\rho_0, \rho_{1N}, \rho_{1S})$ for the genie-aided case is given by

$$\begin{cases} (\text{SNR}_2, 0, \text{SNR}_2) & \text{if } \text{SNR}_2 \leq \text{INR}_C \\ (\frac{\text{SNR}_2 - \beta \text{INR}_C}{\beta}, 0, \text{INR}_C) & \text{if } \text{INR}_C \leq \text{SNR}_2 \end{cases}, \quad (2)$$

On the other hand, when $\text{SIC}_C < \min(\text{SNR}_2, \text{INR}_C)$, then the optimal power allocation $(\rho_0, \rho_{1N}, \rho_{1S})$ is given by,

$$\begin{cases} (\frac{\gamma}{\beta}, 0, \text{SIC}_C) & \text{if } \text{SIC}_C \leq \text{SNR}_2 \leq \text{SIC}'_C \\ (\gamma + \beta\alpha, \gamma - \bar{\beta}\alpha, \text{SIC}_C) & \text{if } \text{SIC}'_C \leq \text{SNR}_2 \leq \text{SIC}'_C + \text{INR}_C - \text{SIC}_C, \\ (\frac{\gamma - \beta\delta}{\beta}, \delta, \text{SIC}_C) & \text{if } \text{SNR}_2 \geq \text{SIC}'_C + \text{INR}_C - \text{SIC}_C \end{cases}, \quad (3)$$

where $\text{SIC}'_C = \text{SIC}_C + \beta \text{INR}_2$, $\alpha = \text{SIC}_C + \text{INR}_2$, $\gamma = \text{SNR}_2 - \beta \text{SIC}_C$ and $\delta = \text{INR}_C - \text{SIC}_C$. Here, β and $\bar{\beta} = 1 - \beta$ are the on and off fractions of the primary, respectively.

Proof: The proof is based on two steps. First we use Lemma 1 to show that Gaussian codebooks are optimal for the secondary transmissions. Then the optimal power allocation is

obtained by solving the following optimization problem,

$$\begin{aligned}
& \max_{\rho_0, \rho_{1N}, \rho_{1S} \geq 0} \quad \bar{\beta} \mathcal{C}(\rho_0) + \beta \left[\mathcal{C} \left(\frac{\rho_{1N}}{1 + \text{INR}_2 + \rho_{1S}} \right) + \mathcal{C}(\rho_{1S}) \right] \\
& \text{subject to} \quad \bar{\beta} \rho_0 + \beta(\rho_{1N} + \rho_{1S}) \leq \text{SNR}_2 \\
& \quad \rho_{1N} + \rho_{1S} \leq \frac{\text{INR}_{\text{gap}}}{|h_{21}|^2} \\
& \quad \rho_{1S} \leq |h_{12}|^2(1 + \text{INR}_{\text{gap}}) - 1.
\end{aligned} \tag{4}$$

The rate achieved is a corner point of the 3 user virtual MAC formed by the primary and the two code layers of the secondary. Hence the sequential decoding at the secondary receiver achieves capacity. Complete details are in Appendix A ■

As shown in Figure 4, power allocation can be viewed as water-filling with a layer in the middle representing the active primary. For the time-slots when primary is off, the secondary uses a single-user codebook matched to its own channel capacity. When the primary is on, the secondary transmitter splits into a two-layer Gaussian superposition codebook as shown in Figure 4 followed by a sequential decoding at the secondary receiver. The sequence of decoding at the secondary receiver is as follows. First, Layer 1 is decoded by treating everything else as noise, then the primary codeword is decoded by treating Layer 2 as noise and finally Layer 2 is decoded interference-free.

B. The General Case: Secondary with noisy estimate of primary state \hat{s}_0

When the secondary does not know the channel states s_t exactly for each time slot, it senses the channel in the beginning of each block \hat{s}_0 (noisy estimate of s_0). The effective channel is a Gaussian mixture channel as given by Equation (1). A given time-slot (for the same sensed state) has the same channel statistics across different blocks as shown in Figure 3. We take a cue from the form of Genie-aided code in Theorem 1 and use a general superposition code. If $\hat{s}_0 = 0$, the secondary transmitter sends at a power level $\boldsymbol{\rho}_0 = \boldsymbol{\rho}_{0N} + \boldsymbol{\rho}_{0S}$, (where $\boldsymbol{\rho}_{0S} = (\rho_{0S1}, \dots, \rho_{0ST})$ etc.) and if $\hat{s}_0 = 1$, it sends at a power level $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_{1N} + \boldsymbol{\rho}_{1S}$. The above time-dependent power allocation exploits secondary's knowledge of the conditional probability distribution of effective noise in each time slot caused by the primary. We shall omit the subscript 0 in s_0, \hat{s}_0 in the summation indices to avoid clutter. The error in the state estimate is characterized by the probability of missed detection $\mathcal{P}_M = P(\hat{s}_0 = 0 | s_0 = 1)$ and the probability of false alarm

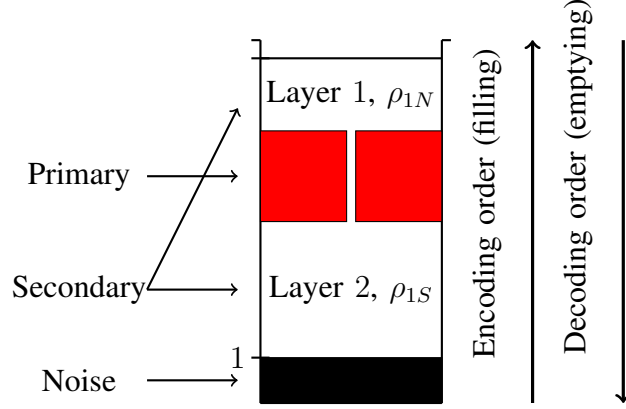


Fig. 4. Layered water-filling: power allocation can be thought of as filling a layered bucket with water. The presence of the primary causes one of the layers to be fixed. The size of the fixed layer (in red) is a function of INR_2 , the size of Layer 2 depends on the interference gap INR_{gap} and the rest of the power is put into Layer 1. When the primary is persistent, there is only one bucket to fill, as shown here. For a sporadic primary, there is another bucket without any layers for ρ_0 . For the block activity model, each of the time slots across the block look alike with (ρ_N, ρ_S) replaced by (ρ_{0Nt}, ρ_{0St}) or (ρ_{1Nt}, ρ_{1St}) depending on the starting state.

$\mathcal{P}_F = P(\hat{s}_0 = 1 | s_0 = 0)$. The exact capacity with noisy state estimate \hat{s}_0 is unknown. To derive an achievable rate using Gaussian codes for secondary transmissions, we extend the optimization problem in (5) by generalizing the rate and the three constraints for each sensed state.

The average power constraint can be computed by summing the power levels weighted with the appropriate probability of occurrences,

$$\begin{aligned}
 & \frac{\pi(0)(1 - \mathcal{P}_F)}{T + 1} \sum_{t=1}^T (\rho_{0Nt} + \rho_{0St}) + \frac{\pi(0)\mathcal{P}_F}{T + 1} \sum_{t=1}^T (\rho_{1Nt} + \rho_{1St}) \\
 & + \frac{\pi(1)\mathcal{P}_M}{T + 1} \sum_{t=1}^T (\rho_{0Nt} + \rho_{0St}) + \frac{\pi(1)(1 - \mathcal{P}_M)}{T + 1} \sum_{t=1}^T (\rho_{1Nt} + \rho_{1St}) \leq \text{SNR}_2 \\
 \text{i.e. } & \sum_{s \in \{0,1\}} \pi(s) \sum_{\hat{s} \in \{0,1\}} P(\hat{s}|s) \sum_{t=1}^T (\rho_{\hat{s}Nt} + \rho_{\hat{s}St}) \leq (T + 1)\text{SNR}_2. \tag{5}
 \end{aligned}$$

Similarly the rate has to be calculated for the four possible combinations of (s_0, \hat{s}_0) . When the primary user is switched ‘off’ or ‘on,’ the effective secondary channel is AWGN with noise $\mathcal{N}(0, 1)$ or $\mathcal{N}(0, 1 + \text{INR}_2)$ respectively. When there is no primary on the channel, the secondary can support a rate of $\mathcal{C}\left(\frac{\rho_{0Ni}}{1 + \rho_{0Si}}\right) + \mathcal{C}(\rho_{0Si}) = \mathcal{C}(\rho_{0Si} + \rho_{0Ni})$ and in the presence of the primary, the secondary can support a rate of $\mathcal{C}\left(\frac{\rho_{0Ni}}{1 + \text{INR}_2 + \rho_{0Si}}\right) + \mathcal{C}(\rho_{0Si})$.

a) $s_0 = 0, \hat{s}_0 = 0$: When $s_0 = 0, \hat{s}_0 = 0$, the secondary correctly detects that the primary is silent and sends $\boldsymbol{\rho}_0 = (\boldsymbol{\rho}_{0N}, \boldsymbol{\rho}_{0S})$. The achievable rate averaged over all possible primary switches is given by,

$$\begin{aligned} R_{00} &= \sum_{t=1}^T \left[f_T(t) \left(\sum_{i=1}^{t-1} \mathcal{C}(\rho_{0Si} + \rho_{0Ni}) + \sum_{i=t}^T \left(\mathcal{C} \left(\frac{\rho_{0Ni}}{1 + \text{INR}_2 + \rho_{0Si}} \right) + \mathcal{C}(\rho_{0Si}) \right) \right) \right] \\ &\stackrel{(a)}{=} \sum_{t=1}^T \left[(1 - F_T(t)) \mathcal{C}(\rho_{0St} + \rho_{0Nt}) + F_T(t) \left(\mathcal{C} \left(\frac{\rho_{0Nt}}{1 + \text{INR}_2 + \rho_{0St}} \right) + \mathcal{C}(\rho_{0St}) \right) \right] \\ &= \sum_{t=1}^T (\bar{\beta}_0(t) \mathcal{C}_{s0} + \beta_0(t) \mathcal{C}_{s1}), \end{aligned}$$

where $\mathcal{C}_{s0} = \mathcal{C}(\rho_{sSt} + \rho_{sNt})$, $\mathcal{C}_{s1} = \mathcal{C} \left(\frac{\rho_{sNt}}{1 + \text{INR}_2 + \rho_{sSt}} \right) + \mathcal{C}(\rho_{sSt})$, $\bar{\beta}_0(t) = 1 - F_T(t)$ and $\beta_0(t) = F_T(t)$. Lemma 3 (see Appendix B) was used to change the summation order in step (a).

b) $s_0 = 1, \hat{s}_0 = 1$: In this case, as the secondary correctly detects the primary user and sends $\boldsymbol{\rho}_0 = (\boldsymbol{\rho}_{0N}, \boldsymbol{\rho}_{0S})$ and the effective noise at the secondary receiver has a variance of $1 + \text{INR}_2$ in the beginning of the block which changes to 1 when the primary user changes its state at time-slot $\tau = t$. Probability of this event $s = 1$, is $\pi(1)$. The rate that can be achieved in this case is given by, $R_{11} = \sum_{t=1}^T (\beta_1(t) \mathcal{C}_{s1} + \bar{\beta}_1(t) \mathcal{C}_{s0})$, where $\bar{\beta}_1(t) = F_T(t)$ and $\beta_1(t) = 1 - F_T(t)$.

c) $s_0 = 1, \hat{s}_0 = 0$: Even though the primary user is transmitting, the secondary detects that there is no primary packet on the channel, it sends $(\boldsymbol{\rho}_{0N}, \boldsymbol{\rho}_{0S})$. In this case the primary user is actually on during the start of the block, the noise that the secondary receiver sees has a variance of $1 + |h_{12}|^2 \text{SNR}_1$ and it changes to 1 when the primary changes its state. Probability of the event $(s_0 = 1, \hat{s}_0 = 0)$, is $\pi(1) \mathcal{P}_M$. The rate achieved for a given t this case is given by, $R_{10} = \sum_{t=1}^T (\beta_1(t) \mathcal{C}_{s1} + \bar{\beta}_1(t) \mathcal{C}_{s0})$.

d) $s_0 = 0, \hat{s}_0 = 1$: In this case, as the secondary detects the primary user and it sends $(\rho_{11}, \dots, \rho_{1T})$, but the primary user is actually off during the start of the block, so the noise that the secondary receiver sees in the beginning is N_2 and it changes to $P_1 + N_2$ when the primary changes its state. Probability of the event $(s_0 = 0, \hat{s}_0 = 1)$ is $\pi(0) \mathcal{P}_F$. The rate achieved for a given t is given by, $R_{01} = \sum_{t=1}^T (\bar{\beta}_0(t) \mathcal{C}_{s0} + \beta_0(t) \mathcal{C}_{s1})$.

Adding up all the terms derived above after weighting them with the appropriate probability of occurrences, leads to the average rate of $R_2(\boldsymbol{\rho}) = \frac{1}{2(T+1)} [\pi(0)(1 - \mathcal{P}_F) R_{00} + \pi(1)(1 - \mathcal{P}_M) R_{11} + \pi(1) \mathcal{P}_M R_{10} + \pi(0) \mathcal{P}_F R_{01}]$.

Two additional constraints are required to complete the problem formulation. First the *INR constraint* which is imposed due to the constraint on the maximum noise variance that can be experienced at the primary receiver, $\rho_{\hat{s}Nt} + \rho_{\hat{s}St} \leq \text{INR}_C$, for $t \in [1, T]$ and $s \in \{0, 1\}$. Second, the *SIC constraint* which is imposed to ensure that the primary information can be decoded in the presence of the multiuser codeword $\rho_{\hat{s}St}$, after the single user codeword $\rho_{\hat{s}Nt}$ has been decoded out. This gives rise to the condition, $\mathcal{C}\left(\frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}}\right) \leq \mathcal{C}\left(\frac{\text{INR}_2}{1+\rho_{sSt}}\right)$ which is same as $\rho_{sSt} \leq \text{SIC}_C$, for $t \in [1, T]$ and $s \in \{0, 1\}$. Note that these two constraints hold for each state and time slot in contrast to the average power constraint which are an average constraint. We solve the following optimization problem, to find the optimal power profile,

$$\max_{\rho \in \Phi} R_2(\rho) = \frac{1}{2(T+1)} \sum_s \pi(s) \sum_{\hat{s}} P(\hat{s}|s) \sum_{t=1}^T (\bar{\beta}_s(t) \mathcal{C}_{\hat{s}0} + \beta_s(t) \mathcal{C}_{\hat{s}1}), \quad (6)$$

where Φ is the constraint set defined by the INR, SIC and average power constraints derived above. Additionally, we have to consider positivity constraints for all the power variables, i.e. $\rho_{\hat{s}Nt}, \rho_{\hat{s}St} \geq 0$ for $\hat{s} \in \{0, 1\}, t \in [1, T]$. This is with the understanding that the constraints do not become infeasible, i.e. $T, \text{INR}_{\text{gap}}, \text{SNR}_2 \geq 0$ and if $|h_{12}|^2(\text{INR}_{\text{gap}} + 1) \leq 1$, $\rho_{sSt} = 0$. We will assume complementarity and positivity but not discuss it further due to lack of space.

Theorem 2 (Monotonicity of the power profile): For the optimization problem given in (6), $\rho_{0N1}^* \geq \rho_{0N2}^* \geq \dots \geq \rho_{0NT}^*$ and $\rho_{1N1}^* \leq \rho_{1N2}^* \leq \dots \leq \rho_{1NT}^*$. Additionally,

- (i) If $\rho_{0Nt}^* = 0$ then $\rho_{0Nt+1}^*, \dots, \rho_{0NT}^* = 0$
- (ii) If $\rho_{1Nt}^* = 0$ then $\rho_{1N1}^*, \dots, \rho_{1Nt-1}^* = 0$
- (iii) If $\rho_{0Nt}^* = \text{INR}_c$ then $\rho_{0N1}^*, \dots, \rho_{0Nt-1}^* = \text{INR}_c$
- (iv) If $\rho_{1Nt}^* = \text{INR}_c$ then $\rho_{1Nt+1}^*, \dots, \rho_{1NT}^* = \text{INR}_c$

Proof: See Appendix C. ■

Due to the generality of our problem formulation in (6), the exact form of the power distribution cannot be derived in closed form. However, Theorem 2 proves a very important result about the monotonicity of power allocation across time-slots. The optimal power profile is non-increasing in time if the start state is $\hat{s}_0 = 0$. That is the secondary gets *paranoid* over time since it does not know when the primary transmitter will start transmitting. So it is better for the secondary to send more power in the initial time-slots and become more conservative as time progresses. In contrast, if $\hat{s}_0 = 1$, then the secondary bets more power in the later time-slots as there is a higher chance that the primary will turn off in those slots, thereby creating a better channel

for the secondary flow. In the next section, we show that the above result can be significantly strengthened for the special case of perfect state estimate $\hat{s}_0 = s_0$.

C. Special Case I: Secondary with perfect estimate of primary state, $\hat{s}_0 = s_0$

In this section we assume that there is no error in secondary's estimate of s_0 . When the starting state is perfectly known, each of the T subchannels shown in Figure 3 behave as parallel channels. For such a channel capacity can be achieved by sending at a constant power with receiver side channel side information [13]. We show below that Gaussian codewords with power levels $\rho_{st} = \rho_{sSt} + \rho_{sNt}$ achieves the capacity if the sensing is error-free.

Theorem 3 (Capacity with perfect sensing, $\hat{s}_0 = s_0$): With perfect sensing, the capacity for the discrete cognitive interference channel is given by,

$$C = \sum_{s_0 \in S} \frac{\pi(s_0)}{T+1} \sum_{t=1}^T \max_{p(X_2^{(t)}|s_0)} \sum_{s \in S} p(s|s_0) I(X_2^{(t)}; Y_2^{(t)}|s_0, s).$$

Proof: See Appendix D. ■

The above result for finite input-output alphabets extends to continuous alphabets such that $I(X_2^{(t)}; Y_2^{(t)}|s_0, s) = \beta_s(t)\mathcal{C}_{s0} + \bar{\beta}_s(t)\mathcal{C}_{s1}$, $\sum_{s_0 \in S} \sum_{t=1}^T \pi(s_0)\rho_{st} \leq \text{SNR}_2$, $\rho_{st} \leq \frac{\text{INR}_{\text{gap}}}{|h_{21}|^2}$ for $t = 1, \dots, T$ and $p(X_2^{(t)}|s_0) \sim \mathcal{N}(0, \rho_{st})$. In order to find the power allocation ρ_{st} , we have to solve an optimization problem similar to (5). When sensing is perfect $\hat{s}_0 = s_0$ and $P(\hat{s}_i|s_0) = 1$. There is no missed detection or false alarm, i.e. $\mathcal{P}_M = 0$ and $\mathcal{P}_F = 0$. So, the rate and all the constraint equations for the perfect sensing protocol can be obtained by making these substitutions in (6).

Theorem 4 (Optimal Layer 2 power): For the optimization problem given in (6) with $\mathcal{P}_M = \mathcal{P}_F = 0$ and $\hat{s}_0 = s_0$, the optimal power of the Layer 2 Gaussian codeword is given by, $\rho_{sSt}^* = (\min(\text{SIC}_C, \text{INR}_C, \text{SNR}_2))^+$, where $\text{SIC}_C, \text{INR}_C$ are defined in Theorem 1 and $x^+ = \max(x, 0)$.

Proof: See Appendix E. ■

Due to the perfect estimate of the primary starting state, the rate splitting done at secondary is optimal, which allows us to find the exact value of the Layer 2 codewords. The monotonicity properties of the power levels still hold for the perfect sensing case.

Theorem 5 (Monotonic Layer 1 profile): For the optimization problem given in (6) with $\mathcal{P}_M = \mathcal{P}_F = 0$, $\rho_{0N1}^* \geq \rho_{0N2}^* \geq \dots \geq \rho_{0NT}^*$ and $\rho_{1N1}^* \leq \rho_{1N2}^* \leq \dots \leq \rho_{1NT}^*$. Additionally,

- (i) If $\rho_{0Nt}^* = 0$ then $\rho_{0Nt+1}^*, \dots, \rho_{0NT}^* = 0$
- (ii) If $\rho_{1Nt}^* = 0$ then $\rho_{1N1}^*, \dots, \rho_{1Nt-1}^* = 0$
- (iii) If $\rho_{0Nt}^* = \text{INR}_c$ then $\rho_{0N1}^*, \dots, \rho_{0Nt-1}^* = \text{INR}_c$
- (iv) If $\rho_{1Nt}^* = \text{INR}_c$ then $\rho_{1Nt+1}^*, \dots, \rho_{1NT}^* = \text{INR}_c$

Proof: The proof follows directly from Theorem 2 for $\hat{s}_0 = s_0$ and $\mathcal{P}_M = \mathcal{P}_F = 0$. ■

The intuition for the above results is as follows. It is always better to allocate more power to Layer 2 codewords (while satisfying the power and interference constraints) as it has a higher contribution towards the secondary rate R_2 . For a fixed $|h_{12}|$, below a certain INR_{gap} , this upper bound is zero and all the power goes to the Layer 1 codeword. As INR_{gap} increases, the proportion of the Layer 2 codeword keeps increasing and in the end all the power is put into the Layer 2 codeword. In short, we have to do layered water-filling for each time slot in the block as shown in Figure 4. Next, we show that the opportunistic superposition of [19] is a special case of the cognitive protocol when no sensing is done.

D. Special Case II: Secondary with no information about primary state s_0

Consider the special case when the secondary does not sense the channel at all, i.e. the secondary transmitter only knows the statistics of the primary traffic $\pi(s_0)$ and $f_T(\tau)$. Alternately, the estimate \hat{s}_0 is so noisy that it does not provide any information about s_0 . A similar analysis of this special case can also be found in [20]. Out of the available power SNR_2 , $\rho_S = \alpha \text{SNR}_2$ is assigned to the Layer 2 codeword and $\rho_N = (1 - \alpha) \text{SNR}_2$ is assigned to the Layer 1 codeword. The average power constraint for the secondary is given by $\rho_N + \rho_S \leq \text{SNR}_2$ and the INR constraint can be written as $\rho_N + \rho_S \leq \frac{\text{INR}_{\text{gap}}}{|h_{21}|^2}$. After decoding the single user codeword, the residual capacity of the channel is given by $\mathcal{C}\left(\frac{\text{INR}_2}{1 + \alpha \text{SNR}_2}\right)$. To ensure primary is decodability the SIC constraint is given by, $R_1 \leq \mathcal{C}\left(\frac{\text{INR}_2}{1 + \alpha \text{SNR}_2}\right)$ (i.e. $\alpha \leq \frac{|h_{12}|^2(\text{INR}_{\text{gap}} + 1) - 1}{\text{SNR}_2}$). For such an α , a secondary rate of $R_2(\alpha) = \mathcal{C}\left(\frac{(1 - \alpha)\text{SNR}_2}{1 + \text{INR}_2 + \alpha \text{SNR}_2}\right) + \mathcal{C}(\alpha \text{SNR}_2)$ can be achieved. Gaussian codebooks are optimal in this case too by Lemma 1. The optimal value of the superposition fraction (α^*) is one which maximizes $R_2(\alpha)$ while satisfying all the constraints.

Theorem 6 (No-sensing Capacity): For a cognitive interference channel where the secondary

does not sense the channel, the optimal superposition fraction is given by

$$\alpha^* = \begin{cases} 0 & \text{if } \text{INR}_2 \leq \frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}} \\ \frac{\text{SIC}_C}{\text{SNR}_2} \frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}} < \text{INR}_2 < \frac{\text{SNR}_1(1+\text{SNR}_2)}{1+\text{INR}_{\text{gap}}} , \\ 1 & \text{if } \text{INR}_2 \geq \frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}}(1 + \text{SNR}_2) \end{cases}$$

and the capacity is given by $R_2(\alpha^*)$.

Proof: See Appendix F. ■

From the secondary receiver's perspective, the equivalent channel is a MAC. The fixed primary converts the equivalent MAC rate region (pentagon) to a single line (Figure 5(a)). Even though the rate splitting assumes sequential decoding, it turns out to be optimal [3] because the two code layers at the secondary, makes the rate tuple a corner point of a three user virtual MAC [20] consisting of the two code layers of secondary and the primary.

Remark 1 (Persistent Primary): If the primary has persistent data ($\pi(1) = 1$), the rate that the secondary can achieve is the same as proved above, i.e. the effective channel in the no-sensing case is a compound channel and the secondary has to code for the worst case.

The optimal distribution puts as much available power in ρ_{1S} as possible without violating the SIC, INR and power constraints. This can be thought of a layered water-filling as shown in Figure 4. The layer due to the primary codeword is fixed. Available power is first assigned to the Layer 2 codeword. When the fixed layer is very high so that no power is left to put in the Layer 1 codeword, $\alpha^* = 0$ and if the fixed layer touches the noise floor, all the power is put in the Layer 1 codeword, $\alpha^* = 1$. The idea of opportunistically doing interference cancellation has also been analyzed in [?, 19] and is a restatement of the rate splitting approach introduced in [20] for achieving time-sharing without cooperating encoders. The difference from [19] is the effect of the secondary transmissions on the primary which gives rise to the power profile and the proof of optimality using rate splitting and sequential decoding.

Figure 5(b) shows a detailed view of the the different rate splitting regions as a function of INR_{gap} and INR_2 , for fixed SNR_1 , SNR_2 . In Region 1, the secondary user treats the primary data as noise ($\alpha^* = 0$). In region 2 the secondary uses both layers. In Region 3 the secondary receiver first decodes the primary's data and then decodes its own ($\alpha^* = 1$). The Layer 1 protocol region is always the same w.r.t. INR_1 . The increase in the Layer 2 protocol region with the increase in the $|h_{21}|^2$ can be attributed to the interference constraint at the primary receiver. This increase

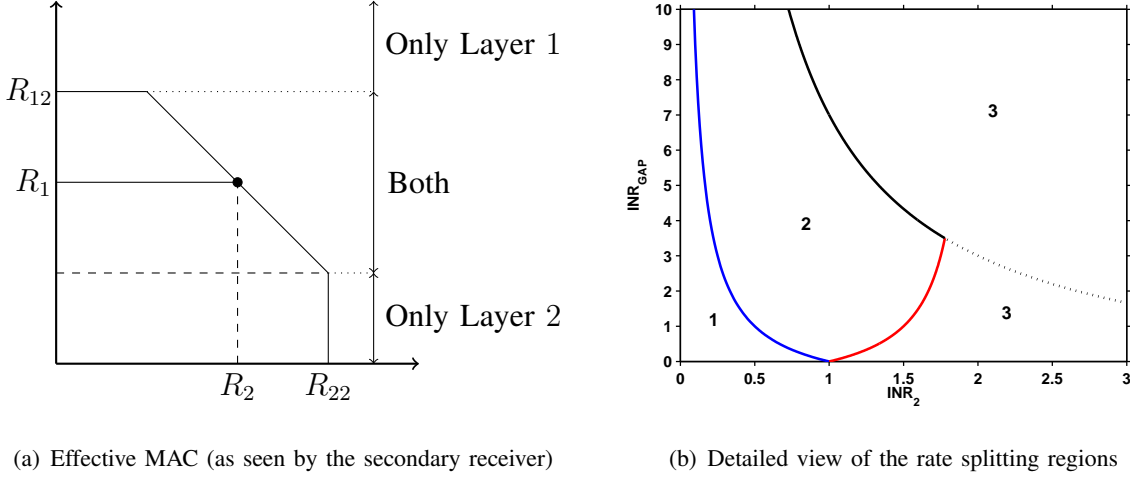


Fig. 5. (A) The effective MAC as seen by the secondary receiver, where $R_1 = C(\frac{INR_2}{1+\alpha SNR_2})$ is the decodable primary rate as seen by the secondary receiver, $R_{12} = C(INR_2)$ is the rate supported by the cross channel between the primary transmitter and the secondary receiver, $R_2(\alpha)$ is the rate achievable by the secondary and $R_{22} = C(SNR_2)$ is the maximum rate achievable by the secondary if there is no primary. (B) Detailed view of the regions of operation for the secondary user for $SNR_1 = SNR_2 = 7$ as a function of the two variables INR_{gap} and INR_2 . The secondary user uses Layer 1 codewords in Region 1, Layer 2 codewords in Region 3 and a superposition code with both layers in Region 2. Note that $INR_{gap} = 0$ is not a part of any region.

comes at the expense of the superposition protocol region. When $|h_{21}|^2$ increases, the power available to the secondary user keeps on decreasing due to the interference constraint. This decreases the power left for the Layer 1 codeword. Hence at higher $|h_{21}|^2$, a part of the mixed protocol region gets converted to a multiuser protocol region as there is no more power left to put into the single user codeword.

The paranoid power profile is optimal if the sensing is perfect. However, even if the sensing is noise free, the time spent in sensing is still an overhead. This overhead can be high enough for the no-sensing scheme to outperform the perfect sensing scheme in some regimes, as shown in Figure 6. Along the curve, points close to the right side ($R_2 = 0$) are obtained for $INR_{gap} = 0$ and the points to the left side ($R_1 = 0$) are obtained for $INR_{gap} = \infty$. Figure 7 plots the two extreme rate points on the y-axis in Figure 6 (where $R_1 = 0$) for different block sizes. The above-mentioned loss due to sensing, decreases as T increases. Finally, the bounds are closer together when the residual capacity of the primary channel is smaller as shown in Figure 8. Different values of β has no effect on the no-information lower bound. But the genie-aided upper bound is able to get a higher rate if the channel is idle for a longer period of time. This

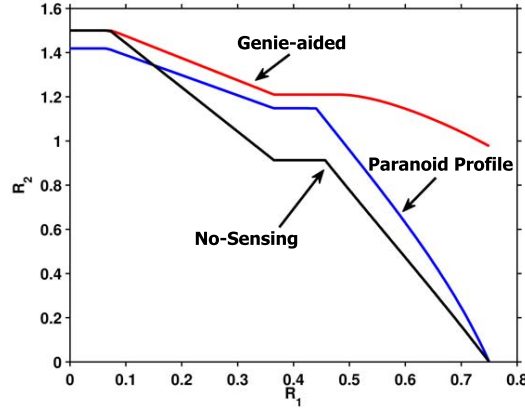


Fig. 6. The no-information lower bound can outperform perfect sensing scheme with a finite sensing overhead depending on the INR_{gap} , the blocklength T . Here, $P_1 = P_2 = 7$, $\beta = 0.5$, $\pi(0) = \pi(1) = 0.5$, $|h_{12}|^2 = |h_{21}|^2 = 0.5$ and $T = 10$.

means, when β is high, the performance of both these bounds are close.

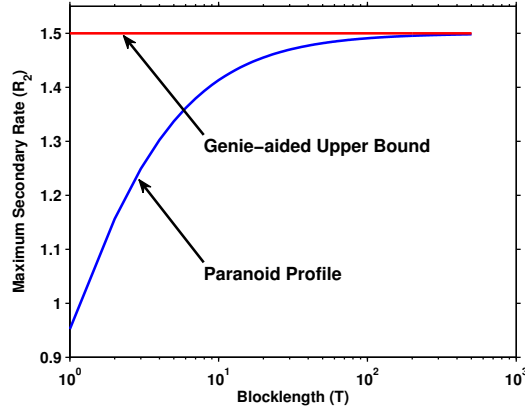


Fig. 7. The maximum R_2 achieved (two extreme rate points on the y-axis in Figure 6 where $R_1 = 0$) by the paranoid profile scheme gets closer to the genie-aided upper bound as T increases which is due to smaller sensing overhead. Here, $P_1 = P_2 = 7$, $\beta = 0.5$, $\pi(0) = \pi(1) = 0.5$, $|h_{12}|^2 = |h_{21}|^2 = 0.5$.

IV. CONCLUSION

We approximate an interference channel in the presence of a fixed sporadic primary flow as a block activity channel. Such a block activity model can be broken into parallel channels in time and multiplexed codebooks can be used when there is perfect sensing. We derived a

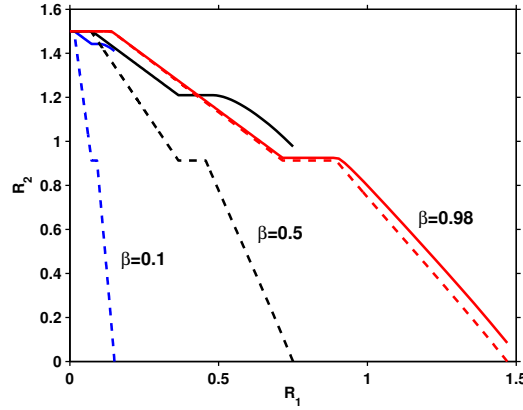


Fig. 8. Comparison of no-sensing scheme and the Genie-aided scheme for different persistence (β) of the primary user's data and $|h_{12}|^2 = |h_{21}|^2 = 0.5$. The dashed lines are for the no-sensing scheme and the solid lines are for the genie-aided scheme.

paranoid scheme for the secondary user when the primary user is transmitting below capacity and show that rate splitting at the secondary transmitter with sequential decoding is optimal when the estimation of the primary starting state is noise-free. Depending on the starting state of the primary during a block, the optimal power profile for Gaussian inputs is either growing or decaying in power as a function of the time slot. If the sensing overhead is considered, we showed that the paranoid scheme approaches the genie-aided scheme for large block lengths. Finally we show numerically that the paranoid and genie aided schemes approach the no-information scheme when the primary channel is operating close to its capacity.

APPENDIX A

Proof of Theorem 1: We derive a special case of the water-filling result [8] for a channel with two states, which will be used to prove the optimal power allocation for the genie-aided case.

Lemma 2: For the optimization problem given by

$$\begin{aligned} \max_{\rho_0, \rho_1 \geq 0} \quad & \bar{\beta}C(\rho_0) + \beta C\left(\frac{\rho_1}{1+\alpha}\right) \\ \text{subject to} \quad & \bar{\beta}\rho_0 + \beta\rho_1 \leq \gamma \text{ and } \rho_1 \leq \delta \end{aligned} \quad (7)$$

with non-negative parameters $(\alpha, \beta, \gamma, \delta)$, the maximizing ρ_0, ρ_1 is characterized as follows.

$$(\rho_0, \rho_1) = \begin{cases} (\frac{\gamma}{\bar{\beta}}, 0) & \text{if } \rho_1^* < 0 \leq \delta & (A) \\ (\rho_0^*, \rho_1^*) & \text{if } 0 \leq \rho_1^* \leq \delta & (B) \\ (\frac{\gamma - \beta\delta}{\bar{\beta}}, \delta) & \text{if } 0 \leq \delta < \rho_1^* & (C) \end{cases}$$

where $\rho_0^* = \gamma + \beta\alpha$, $\rho_1^* = \gamma - \bar{\beta}\alpha$ and $\bar{\beta} = 1 - \beta$.

Proof: Let us start with assuming that there is no constraint on ρ_1 . The optimal solution is given by $\rho_0 = (\frac{1}{\lambda} - 1)^+$, $\rho_1 = (\frac{1}{\lambda} - 1 - \alpha)^+$, where λ is chosen such that $\bar{\beta}\rho_0 + \beta\rho_1 = \gamma$. If $\frac{1}{\lambda} > 1 + \alpha$, then $\frac{1}{\lambda} = \gamma + \bar{\beta} + \beta(1 + \alpha)$. Therefore, $\rho_0^* = \gamma + \beta\alpha$ and $\rho_1^* = \gamma - \bar{\beta}\alpha$. Note $\frac{1}{\lambda} > 1 + \alpha$ also ensures $\rho_1^* \geq 0$ and by assumption $\rho_1^* \leq \delta$. This proves part B of the Lemma.

If $\frac{1}{\lambda} \leq 1 + \alpha$, $\rho_1 = 0$ and λ is such that $\bar{\beta}\rho_0 = \gamma$, then $\rho_0 = \frac{\gamma}{\bar{\beta}}$, $\rho_1 = 0$. This proves part A of the solution. As long as ρ_1^* lies below δ (i.e. the extra constraint on ρ_1 is not active at the solution), the same solutions hold. If $\rho_1^* > \delta$, $\rho_1 = \delta$, $\rho_0 = \frac{\gamma - \beta\delta}{\bar{\beta}}$. ■

The constraint set defined by the average power constraint at the secondary transmitter, the INR constraint at the primary receiver, and the SIC constraint at the secondary receiver is schematically shown by the three colored planes in Figure 9.

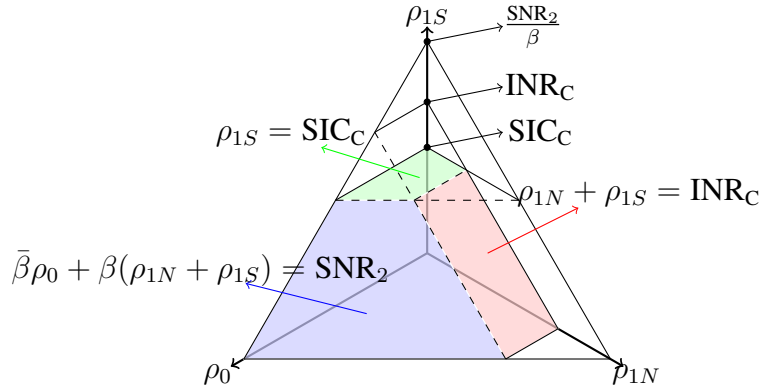


Fig. 9. General Constraint set for the perfect sensing optimization problem.

First consider the case when the SIC constraint is not active i.e. when $\text{SIC}_C \geq \min(\text{SNR}_2, \text{INR}_C)$. In this case, the solution lies on the $\rho_0 - \rho_{1S}$ plane of Figure 9. This is because $R_2(\rho_0, \rho_{1N}, \rho_{1S}) \leq R_2(\rho_0, \rho_{1N} - \kappa, \rho_{1S} + \kappa)$ as long as $\kappa \leq \rho_{1N}$, i.e. we can achieve a higher rate by taking away some power from ρ_{1N} and giving it to ρ_{1S} as long as all the constraints are satisfied. Therefore, to maximize the rate, all the available power is allocated to ρ_{1S} since there is no SIC constraint.

When we put $\rho_{1N} = 0$, the optimization problem looks like Equation (7) with $\alpha = 0, \gamma = \text{SNR}_2$ and $\delta = \text{INR}_C$. By using Lemma 2, the solution is given by Equation 2.

Finally, if $\text{SIC}_C < \min(\text{SNR}_2, \text{INR}_C)$, the SIC constraint is active at the solution and the solution lies on the SIC constraint plane of Figure 9. In contrast to the above cases, we cannot put all the power from ρ_{1N} in ρ_{1S} . We increase ρ_{1S} till $\rho_{1S} = \text{SIC}_C$ which equals the SIC constraint. Now the problem is same as Equation 7 with $\alpha = |h_{12}|^2(1 + \text{INR}_{\text{gap}} + \text{SNR}_1) - 1, \gamma = \text{SNR}_2 - \beta \text{SIC}_C$ and $\delta = \text{INR}_C - \text{SIC}_C$. Using Lemma 2, and the solution is given by Equation 3.

APPENDIX B

Lemma 3 (Change in Order of Summation): Given a probability mass function, $f(t), 1 \leq t \leq T$ of a discrete random variable, and positive real numbers $(\rho_1, \dots, \rho_T), \alpha$ and β ,

$$\sum_{j=1}^{T+1} f(j) \left[\sum_{i=1}^{j-1} C\left(\frac{\rho_i}{\alpha}\right) + \sum_{i=j}^T C\left(\frac{\rho_i}{\beta}\right) \right] = \sum_{i=1}^T \left[(1 - F(t))C\left(\frac{\rho_i}{\alpha}\right) + F(t)C\left(\frac{\rho_i}{\beta}\right) \right],$$

where $C(x) = \log(1 + x)$, $F(j) = \sum_{i=1}^j f(i), 1 \leq j \leq T$ and $F(0) = 0, F(T+1) = 1$.

Proof: Let us first consider the first term of the left hand side with $\alpha = 1$,

$$\begin{aligned} \sum_{j=1}^{T+1} f(j) \sum_{i=1}^{j-1} c(\rho_i) &= f(2)c(\rho_1) + f(3)[c(\rho_1) + c(\rho_2)] + \dots + f(T+1)[c(\rho_1) + \dots + c(\rho_T)] \\ &= c(\rho_1)[f(2) + \dots + f(T+1)] + \dots + c(\rho_T)f(T+1) = \sum_{i=1}^T c(\rho_i) \sum_{j=i+1}^{T+1} f(j). \end{aligned}$$

From the definition of the probability distribution function, $1 - F(j) = \sum_{i=1}^{T+1} f(i) - \sum_{i=1}^j f(i) = \sum_{i=j+1}^{T+1} f(i)$. Hence, $\sum_{j=1}^{T+1} f(j) \sum_{i=1}^{j-1} c(\rho_i) = \sum_{i=1}^T c(\rho_i) \sum_{j=i+1}^{T+1} f(j) = \sum_{i=1}^T (1 - F(i))c(\rho_i)$. Similarly, $\sum_{j=1}^{T+1} f(j) \sum_{i=t}^T c(\rho_i) = \sum_{i=1}^T F(t)c(\rho_i)$, which completes the proof. ■

APPENDIX C

Proof of Theorem 2: We will drop the subscript 0 on s, \hat{s} to reduce clutter. The Lagrangian for the optimization problem in Theorem 2 can be written as,

$$L(\boldsymbol{\rho}, \lambda_1, \lambda_2) = R_2(\boldsymbol{\rho}) - \lambda_1 g_1(\boldsymbol{\rho}) - \sum_{s,t} \lambda_{2\hat{s}t} g_{2\hat{s}t}(\boldsymbol{\rho}) - \sum_{s,t} \lambda_{3\hat{s}t} g_{3\hat{s}t}(\boldsymbol{\rho}) - \sum_{s,t} \lambda_{4\hat{s}t} \rho_{\hat{s}Nt} - \sum_{s,t} \lambda_{5\hat{s}t} g_{\hat{s}St}.$$

The partial derivatives of the Lagrangian is given by,

$$\begin{aligned}\frac{\delta L}{\delta \rho_{sNt}} &= - \sum_s \pi(s) p(\hat{s}|s) \left(\frac{\bar{\beta}_s(t)}{1 + \rho_{sNt} + \rho_{sSt}} + \frac{\beta_s(t)}{1 + \text{INR}_2 + \rho_{sNt} + \rho_{sSt}} \right) - \lambda_1 \sum_s \pi(s) p(\hat{s}|s) - \lambda_{3st} - \lambda_{4st} \\ \frac{\delta L}{\delta \rho_{sSt}} &= - \sum_s \pi(s) p(\hat{s}|s) \left(\frac{\bar{\beta}_s(t)}{1 + \rho_{sNt} + \rho_{sSt}} + \frac{\beta_s(t)}{(1 + \text{INR}_2 + \rho_{sNt} + \rho_{sSt})(1 + \text{INR}_2 + \rho_{sSt})} \right) \\ &\quad - \lambda_1 \sum_s \pi(s) p(\hat{s}|s) - \lambda_{2st} - \lambda_{3st} - \lambda_{5st}.\end{aligned}$$

The first order necessary conditions (the complementarity conditions are excluded for lack of space) are given by the following $12T + 1$ equations, $\frac{\delta L}{\delta \rho_{sNt}} = 0$, $\frac{\delta L}{\delta \rho_{sSt}} = 0$, $g_1(\rho) \geq 0$, $g_{2st} \geq 0$, $g_{3st} \geq 0$, and $\rho_{sNt}, \rho_{sSt} \geq 0$, for $s \in \{0, 1\}$, $t \in [1, T]$. If $|h_{12}|^2(\text{INR}_{\text{gap}} + 1) < 1$, then $\rho_{sSt} = 0$. Hence g_{2st} , λ_{2st} are redundant, and by complementarity $\lambda_{5st} = 0$. Similarly, if $0 < \rho_{sNt} < \text{INR}_C$ by complementarity $\lambda_{3st} = 0$ and $\lambda_{4st} = 0$. As the objective is an increasing function in each ρ_{sNt} , the average power constraint is met with equality, i.e. $\lambda_1 \geq 0$. For $|h_{12}|^2(1 + \text{SNR}_{\text{gap}}) < 1$ and $0 < \rho_{sSt} < \frac{P_0}{|h_{21}|^2}$, we have

$$\sum_s \pi(s) p(\hat{s}|s) \left(\frac{\bar{\beta}_s(t)}{1 + \rho_{sNt}} + \frac{\beta_s(t)}{1 + \text{INR}_2 + \rho_{sNt}} \right) = \lambda_1 \sum_s \pi(s) p(\hat{s}|s) \text{ for all } s \in \{0, 1\} \text{ and } t \in [1, T].$$

For a given \hat{s} independent of any t , the right hand side of the above equation is constant.

Rewriting the left hand side for $\hat{s} = 0$,

$$\begin{aligned}\pi(0)(1 - P_F) \left(\frac{\bar{\beta}_0(t)}{1 + \rho_{0Nt}} + \frac{\beta_0(t)}{1 + \text{INR}_2 + \rho_{0Nt}} \right) + \pi(1)P_M \left(\frac{\bar{\beta}_1(t)}{1 + \rho_{0Nt}} + \frac{\beta_1(t)}{1 + \text{INR}_2 + \rho_{0Nt}} \right) &= \text{const.} \\ \text{i.e. } \frac{\pi(0)(1 - P_F)(1 - F(t)) + \pi(1)P_M F(t)}{1 + \rho_{0Nt}} + \frac{\pi(0)(1 - P_F)F(t) + \pi(1)P_M(1 - F(t))}{1 + \text{INR}_2 + \rho_{0Nt}} &= \text{const.} \\ \text{i.e. } \frac{G(t)}{1 + \rho_{0Nt}} + \frac{H(t)}{1 + \text{INR}_2 + \rho_{0Nt}} &= \text{const.},\end{aligned}$$

where $G(t) = \pi(0)(1 - P_F)(1 - F(t)) + \pi(1)P_M F(t)$ and $H(t) = \pi(0)(1 - P_F)F(t) + \pi(1)P_M(1 - F(t))$. Rewriting the equation again,

$$\underbrace{\frac{G(t)}{1 + \rho_{0Nt}} + \frac{H(t)}{1 + \text{INR}_2 + \rho_{0Nt}}}_{A_t} = \underbrace{\frac{G(t+1)}{1 + \rho_{0Nt}} + \frac{H(t+1)}{1 + \text{INR}_2 + \rho_{0Nt}}}_{B_t} + C_t,$$

$$\begin{aligned}\text{where } C_t &= \frac{G(t) - G(t+1)}{1 + \rho_{0Nt}} - \frac{H(t+1) - H(t)}{1 + \text{INR}_2 + \rho_{0Nt}} \\ &= (\pi(0)(1 - P_F) - \pi(1)P_M)(F(t+1) - F(t)) \left(\frac{1}{1 + \rho_{0Nt}} - \frac{1}{1 + \text{INR}_2 + \rho_{0Nt}} \right).\end{aligned}$$

To make the right hand side A_{t+1} , we have to remove C_t and replace ρ_{0Nt} by $\rho_{0N(t+1)}$ in B_t . Since $C_t \geq 0$ when $\pi(0)(1 - P_F) \geq \pi(1)P_M$, and B_t is an decreasing function of ρ_{0Nt} , we have $\rho_{0Nt}^* \geq \rho_{0N(t+1)}^*$. Equality holds when $F_T(t) = F_T(t+1)$. As this is true for any $t \in [1, T]$, $\rho_{0N1}^* \geq \rho_{0N2}^* \geq \dots \geq \rho_{0NT}^*$.

For a given s , say $s = 0$, if there is a t such that $\rho_{0Nt}^* = 0$, then we will now prove by contradiction that $\rho_{0Nt+1}^* = \dots = \rho_{0NT}^* = 0$. Let us assume that $\rho_{0Nt+1}^* = \rho > 0$. Define $R(y, z) = R_2(\rho_{0N1}, \dots, \rho_{0Nt} = y, \rho_{0Nt+1} = z, \dots, \rho_{0NT}, \rho_{1N1}, \dots, \rho_{1NT})$. Note $R_2(0, \rho) < R_2(\rho, 0)$ for any $t \in [1, T]$ and any $\rho > 0$ which satisfies all the constraints.

$$\begin{aligned} R_2(\rho, 0) - R_2(0, \rho) &= (1 - F(t))\mathcal{C}(\rho) + F(t)\mathcal{C}\left(\frac{\rho}{1 + \text{INR}_2}\right) \\ &\quad - (1 - F(t+1))\mathcal{C}(\rho) + F(t+1)\mathcal{C}\left(\frac{\rho}{1 + \text{INR}_2}\right) \\ &= (F(t+1) - F(t))\left(\mathcal{C}(\rho) - \mathcal{C}\left(\frac{\rho}{1 + \text{INR}_2}\right)\right) > 0, \end{aligned}$$

since $\text{INR}_2 > 0$ and $F(t+1) \geq F(t)$. Hence if $\rho_{0Nt} = 0$ then $\rho_{0Nt+1} = 0$ and similarly for $t+2, \dots, T$. We can prove the other parts of the theorem in the exact same way. For all previous time slots, the ordering follows the same order,

- If $\rho_{0Nt=0}$ then $\rho_{0N1}^* \geq \rho_{0N2}^* \geq \dots \geq \rho_{0Nt-1}^* > \rho_{0Nt}^* = \dots = \rho_{0NT}^* = 0$.

And due to the same reasons,

- If $\rho_{1Nt} = 0$ then $\rho_{1N1}^* = \rho_{1N2}^* = \dots = \rho_{1Nt-1}^* = 0 \leq \rho_{1Nt}^* \leq \dots \leq \rho_{1NT}^*$.
- If $\rho_{0Nt} = \frac{P_0}{|h_{21}|^2}$ then $\rho_{0N1}^* = \rho_{0N2}^* = \dots = \rho_{0Nt-1}^* = \frac{P_0}{|h_{21}|^2} \geq \rho_{0Nt}^* \geq \dots \geq \rho_{0NT}^*$.
- If $\rho_{1Nt} = \frac{P_0}{|h_{21}|^2}$ then $\rho_{1N1}^* \geq \rho_{1N2}^* \geq \dots \geq \rho_{1Nt-1}^* > \rho_{1Nt}^* = \dots = \rho_{1NT}^* = \frac{P_0}{|h_{21}|^2}$.

Note, for the case when $\text{SIC}_C > 0$ then $\rho_{sSt} > 0$. From Theorem 4, the value of ρ_{sSt} is fixed. Hence, we can remove it from our problem statement along with g_{2st} and λ_{2st} . The $\frac{\delta L}{\delta \rho_{sSt}}$ part of the Lagrangian follows in the same lines.

APPENDIX D

Proof of Theorem 3: The encoding and decoding functions for each time-slot is same as given in Section II-C. The primary state process $\{S_t\}$ is assumed to be an irreducible, aperiodic, finite-state homogeneous Markov chain and is independent of the secondary channel's input and outputs.

Achievability follows from Theorem 1 in [22], when the state is perfectly known. We give an outline here.

1) *Achievability*: During the i^{th} block, with the starting state s_0 , the encoder chooses the i^{th} symbol of the T codewords of length $\pi(s_0)B$ for T messages and multiplexes them $\rho_i(s_0) = (\rho_1(s_0), \dots, \rho_T(s_0))$, $i = 1 \dots, B$. For large B the capacity of the component channels $C_{s_0}(t) = \max_{p(x|s_0)} \sum_s p(s|s_0) I(X; Y|s, s_0)$ can be achieved. The final rate is calculated by summing over the T multiplexed codewords, $C = \sum_{s_0} \frac{\pi(s_0)}{T+1} \sum_{t=1}^T C_{s_0}(t)$ which concludes the achievability.

2) *Converse*: Let W be the message random variable. The capacity can be written as

$$(T+1)C = \limsup_n \frac{1}{n} \sum_{s_0} \pi(s_0) \sum_{t=1}^T \log M \stackrel{(a)}{\leq} \limsup_n \frac{1}{n} \sum_{s_0} \pi(s_0) \max_{\rho_1^{(t)n}} \sum_{t=1}^T I(W^{(t)}; Y_1^{(t)n}, S_0, S_1^{(t)n}),$$

where (a) follows the same steps as in Equation (4) of [22], and the superscript $(t)n$ refers to time slot t of block n . Now,

$$\begin{aligned} \sum_{t=1}^T I(W^{(t)}; Y_1^{(t)n}, S_0, S_1^{(t)n}) &\stackrel{(b)}{=} \sum_{t=1}^T I(W^{(t)}; Y_1^{(t)n} | S_0, S_1^{(t)n}) \\ &= \sum_{t=1}^T \sum_{i=1}^n \left[H(Y_i^{(t)} | S_0^{(i)}, S_1^{(t)n}, Y_1^{(t)i-1}) - H(Y_i^{(t)} | S_0^{(i)}, S_1^{(t)n}, W^{(t)}, Y_1^{(t)i-1}) \right] \\ &\stackrel{(c)}{\leq} \sum_{t=1}^T \sum_{i=1}^n \left[H(Y_i^{(t)} | S_0^{(i)}, S_i^{(t)}) - H(Y_i^{(t)} | S_0^{(i)}, S_i^{(t)}, X_i^{(t)}) \right] \\ &= \sum_{t=1}^T \sum_{i=1}^n I(X_i^{(t)}; Y_i^{(t)} | S_0^{(i)}, S_i^{(t)}), \end{aligned}$$

where (b) follows from the fact that the state is independent of the message and (c) follows from the fact that entropy decreases on conditioning and Y_i is independent of other random variables when conditioned on W_i and $S_0^{(i)}$. Finally,

$$\begin{aligned} C &\leq \limsup_n \sum_{s_0} \frac{\pi(s_0)}{n(T+1)} \sum_{t=1}^T \max_{X_1^{(t)n}} \sum_{i=0}^{n-1} I(X_i^{(t)}; Y_i^{(t)} | S_0^{(i)}, S_i^{(t)}) \\ &\stackrel{(d)}{=} \limsup_n \sum_{s_0} \frac{\pi(s_0)n}{n(T+1)} \sum_{t=1}^T \max_{p(X^{(t)}|S_0)} I(X^{(t)}; Y^{(t)} | S_0, S^{(t)}) \\ &\leq \sum_{s_0} \frac{\pi(s_0)}{T+1} \sum_{t=1}^T \max_{p(X^{(t)}|S_0)} I(X^{(t)}; Y^{(t)} | S_0, S^{(t)}), \end{aligned}$$

where (d) follows the same steps as Equation (8) in [22] which completes the proof.

APPENDIX E

Proof of Theorem 4: The solution can be rewritten as,

$$\rho_{sSt}^* = \begin{cases} 0 & \text{if } \text{SIC}_C < 0 \\ \text{SIC}_C & \text{if } 0 \leq \text{SIC}_C < \min(\text{SNR}_2, \text{INR}_C) \\ \min(\text{SNR}_2, \text{INR}_C) & \text{if } \text{SIC}_C \geq \max(\text{SNR}_2, \text{INR}_C) \end{cases} \quad (8)$$

The first part follows trivially from the definition of the problem. To prove the other parts, let the optimal solution to the optimization problem be given by (ρ_0^*, ρ_1^*) such that $\rho_{sSt}^* < \text{SIC}_C$ for some $t \in [1, T]$ and $s \in \{0, 1\}$. Let us consider $\rho_{sNt}^*, \rho_{sSt}^*$, keeping all other powers fixed. The constraints for this pair (except the positivity and complementarity constraints) are $A + \pi(s)(\rho_{sNt}^* + \rho_{sSt}^*) \leq T\text{SNR}_2$, $\rho_{sSt}^* \leq \text{SIC}_C$, and $\rho_{sNt}^* + \rho_{sSt}^* \leq \text{INR}_C$, where the contribution to the average power from all other powers except $(\rho_{sNt}^*, \rho_{sSt}^*)$ are lumped into A . The rate can be written as,

$$R_2(\rho_{sNt}^*, \rho_{sSt}^*) = R_2^{(0)} + \pi(s) \left[\bar{\beta}_s(t) \mathcal{C}(\rho_{sNt}^* + \rho_{sSt}^*) + \beta_s(t) \left(\mathcal{C}\left(\frac{\rho_{sNt}^*}{1 + \text{INR}_2 + \rho_{sSt}^*}\right) + \mathcal{C}(\rho_{sSt}^*) \right) \right].$$

The contribution due to all other powers to the average rate has been lumped into $R_2^{(0)}$. Now let us take away a small part of the power from ρ_{sNt}^* and add it to ρ_{sSt}^* . This has no effect on the sum $\rho_{sNt}^* + \rho_{sSt}^*$, hence the constraints remain unchanged. We chose a $\delta > 0$ such that $\rho_{sNt}^* - \delta \geq 0$ and $\rho_{sSt}^* + \delta \leq \text{SIC}_C$. It is easy to see that $R_2(\rho_{sNt}^*, \rho_{sSt}^*) < R_2(\rho_{sNt}^* - \delta, \rho_{sSt}^* + \delta)$,

$$\begin{aligned} \text{i.e. } & \mathcal{C}\left(\frac{\rho_{sNt}^*}{1 + \text{INR}_2 + \rho_{sSt}^*}\right) + \mathcal{C}(\rho_{sSt}^*) < \mathcal{C}\left(\frac{\rho_{sNt}^* - \delta}{1 + \text{INR}_2 + \rho_{sSt}^* + \delta}\right) + \mathcal{C}(\rho_{sSt}^* + \delta) \\ \text{i.e. } & \left(1 + \frac{\rho_{sNt}^*}{1 + \text{INR}_2 + \rho_{sSt}^*}\right)(1 + \rho_{sSt}^*) < \left(1 + \frac{\rho_{sNt}^* - \delta}{1 + \text{INR}_2 + \rho_{sSt}^* + \delta}\right)(1 + \rho_{sSt}^* + \delta) \\ \text{i.e. } & \frac{(1 + \text{INR}_2 + \rho_{sNt}^* + \rho_{sSt}^*)(1 + \rho_{sSt}^*)}{1 + \text{INR}_2 + \rho_{sSt}^*} < \frac{(1 + \text{INR}_2 + \rho_{sNt}^* + \rho_{sSt}^*)(1 + \rho_{sSt}^* + \delta)}{1 + \text{INR}_2 + \rho_{sSt}^* + \delta} \end{aligned}$$

since $\text{INR}_2 > 0$. This perturbation can only go on till either ρ_{sNt}^* reaches its lower limit, or ρ_{sSt}^* reaches the upper limit. We hit the lower limit of ρ_{sNt}^* first if $\text{SIC}_C \geq \max(\text{SNR}_2, \text{INR}_C)$. In this case, $\rho_{sNt}^* = 0$ and $\rho_{sSt}^* = \min(\text{SNR}_2, \text{INR}_C)$. Otherwise, we hit the upper limit ρ_{sSt}^* first, if $\rho_{sSt}^* = \text{SIC}_C$. In either case, the result is independent of the particular value of t .

APPENDIX F

Proof of Theorem 6: The only variable is the superposition fraction α , so the problem is one dimensional. The derivative of the optimization function is given by, $\frac{dR_2(\alpha)}{d\alpha} = \frac{\text{INR}_2 \text{SNR}_1 \text{SNR}_2}{(1 + \alpha \text{SNR}_2)(1 + \text{INR}_2 + \alpha \text{SNR}_2)}$.

The slope of the optimization function is always positive, i.e. $\frac{dR_2(\alpha)}{d\alpha} > 0$. Hence, the optimal solution is achieved at the upper boundary of the intersection of the constraints on α which are $0 \leq \alpha \leq 1$ and $\alpha \leq \frac{\text{INR}_2(1+\text{INR}_{\text{gap}})-1}{\text{SNR}_2}$. If the right hand side is less than or equal to 0, then $\alpha^* = 0$, if the right hand side is more than 1, $\alpha^* = 1$, otherwise, $\alpha^* = \frac{\text{INR}_2(1+\text{INR}_{\text{gap}})-1}{\text{SNR}_2}$. For $|h_{12}| = 1$, this is same as the rate splitting result for two users as shown in [20].

a) If $\text{INR}_2 \geq \frac{\text{SNR}_1(1+\text{SNR}_2)}{1+\text{INR}_{\text{gap}}}$: This comes from the decodability condition $\mathcal{C}\left(\frac{\text{INR}_2}{1+\text{SNR}_2}\right) \geq \mathcal{C}\left(\frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}}\right)$. Hence, after decoding off the primary, the rate that can be achieved by the secondary is $R_2(\alpha^*) = \mathcal{C}(\text{SNR}_2)$. But this is the single user capacity of the secondary channel for the given power constraint when there is no primary interference. Therefore, this scheme of using only a Layer 2 codeword is optimal in this case.

b) If $\text{INR}_2 \leq \frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}}$: the primary cannot be decoded even if no secondary data is sent i.e. $\mathcal{C}(\text{SNR}_2) \leq \mathcal{C}\left(\frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}}\right)$. So, in this case, the only option is to treat this undecodable signal as noise. The effective channel now behaves as a point to point channel with AWGN $\mathcal{N}(0, 1+\text{INR}_2)$. Hence the capacity of this channel is given by $R_2(\alpha^*) = \mathcal{C}\left(\frac{\text{SNR}_2}{1+\text{INR}_2}\right)$.

c) If $\frac{\text{SNR}_1(1+\text{SNR}_2)}{1+\text{INR}_{\text{gap}}} < \text{INR}_2 < \frac{\text{SNR}_1}{1+\text{INR}_{\text{gap}}}$: For the rest of the parameter range, the rate is a corner point of the 3 user virtual MAC formed by the primary and the two layers of the secondary (Figure 5(a)). The fixed primary transceiver converts the effective MAC (from the point of view of the secondary receiver) into a straight line and the secondary splits its information into two layers to achieve the boundary point. The sum capacity of the effective MAC is achieved by this scheme since for any α , $\mathcal{C}\left(\frac{(1-\alpha)\text{SNR}_2}{1+\text{INR}_2+\alpha\text{SNR}_2}\right) + \mathcal{C}\left(\frac{\text{INR}_2}{1+\alpha\text{SNR}_2}\right) + \mathcal{C}(\alpha\text{SNR}_2) = \mathcal{C}(\text{INR}_2 + \text{SNR}_2)$. Additionally α^* gives the best rate among all the secondary rates. Hence the layered encoding at the secondary transmitter and sequential decoding at the secondary receiver is optimal and the capacity is given by $R_2(\alpha^*)$.

REFERENCES

- [1] A. B. Carleial, "Interference channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–70, jan 1978.
- [2] M. H. M. Costa, "On the gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 5, pp. 607–615, sept 1985.
- [3] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [4] D. Dash and A. Sabharwal, "Secondary transmission profile for a single-band cognitive interference channel," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, oct 2008.
- [5] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, may 2006.

- [6] K. Eswaran, M. Gastpar, and K. Ramchandran, "Bits through ARQs," *arxiv:0806.1549v1*, 2008. [Online]. Available: <http://arxiv.org/pdf/0806.1549v1>
- [7] Federal Communications Commission Spectrum Policy Task Force, "Report of the spectrum efficiency working group," *Technical Report 02-135*, nov 2002.
- [8] R. G. Gallager, *Information theory and reliable communication*. John Wiley and Sons, 1968.
- [9] M. Gastpar, "On capacity under receive and spatial spectrum-sharing constraints," *IEEE Trans. Inf. Theory*, vol. 53, pp. 471–487, 2007.
- [10] A. Ghasemi and E. S. Sousa, "Capacity of fading channels under spectrum-sharing constraints," in *Proc. IEEE Int. Conf. on Commun., ICC'06*, vol. 10, jun 2006, pp. 4373–4378.
- [11] —, "Fundamental limits of spectrum-sharing in fading environments," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 649–658, feb 2007.
- [12] A. J. Goldsmith, S. A. Jafar, I. maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: an information theoretic perspective," *Proc. of the IEEE*, vol. 97, no. 5, pp. 894–914, may 2009.
- [13] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, nov 1997.
- [14] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, jan 1981.
- [15] S. A. Jafar and S. Srinivasa, "Capacity limits of cognitive radio with distributed and dynamic spectral activity," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 529–537, 2007.
- [16] A. Jovicic and P. Viswanath, "Cognitive radio: an information-theoretic perspective," may 2006, unpublished.
- [17] A. Lapidoth, "Nearest neighbor decoding for additive non-gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, sep 1996.
- [18] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, KTH, Stockholm, Sweden, dec 2000.
- [19] P. Popovski, H. Yomo, K. Nishimori, R. D. Taranto, and R. Prasad, "Opportunistic interference cancellation in cognitive radio systems," in *Proc. IEEE Int. Symp. New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, 2007.
- [20] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, mar 1996.
- [21] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Stable throughput of cognitive radios with and without relaying capability," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 2351–2360, dec 2007.
- [22] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, mar 1999.
- [23] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4391–4399, nov 2007.
- [24] R. Zhang, "On peak versus average interference power constraints for spectrum sharing in cognitive radio networks," *arxiv:0806.0676v12*, 2008. [Online]. Available: <http://arxiv.org/pdf/0806.0676v2>
- [25] W. Zhang and U. Mitra, "A spectrum-shaping perspective on cognitive radio," in *Proc. IEEE Int. Symp. New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, oct 2008, pp. 1–12.